BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

REMARKS

Applicants have amended claim 26 in an attempt to clarify the designation of the ERL referred to. What is specifically defined is the portion of the ERL at the N-terminus of the second module. It is that portion that is represented by SEQ. ID. Nos.: 12-19. The kind suggestion of the Examiner with respect to claims 28-39 has been adopted, thus obviating the rejection as to redundancy. No new matter has been added and entry of the amendment is respectfully requested.

At the outset, applicants wish to express their appreciation to the Examiner, not only for the withdrawal of a goodly number of rejections previously made, but also for the highly professional and clear manner in which the status of the claims has been presented. Applicants are especially appreciative of the summary of the issues, and the clear accounting, one way or the other, of the rejections previously made. Applicants response to the rejections maintained is set forth below.

Formal Matters

Enclosed herewith is Figure 3 as a formal drawing. The objection to the redundancy in claims 28-39 has been obviated by amendment.

The Rejection Under 35 U.S.C. § 112, Paragraph 2

Applicants appreciate the recognition that claim 25 adequately clarifies the RAL, which are the only subjects of that claim. Therefore, it is believed that claim 25 is sufficiently clear. As amended, claim 26 is clear as well. While claim 26 may not define the entire ERL, the portions that are specified are clearly defined.

With regard to the generic issue, applicants do not define a consensus sequence for either the ERL's or RAL's because no such consensus exists, as the Examiner correctly points out.

However, this does not mean that the skilled artisan would not be able to identify which sequences represent these moieties from art-known materials. By examining the consensus sequences of the various modules, either of the same PKS or of different PKS, it is apparent where the boundaries of the linkers must be. Enclosed is a copy of an article by Donadio, *et al.*, *Gene* (1992) 111:51-60. Figure 2 of this article sets forth in "stacked" form the sequences of modules 1-6 of the erythromycin PKS. With respect to the RAL's, which exist between M1 and M2, M3 and M4, and M5 and M6, it will be seen that the ends of the consensus sequences which comprise these modules are clearly marked. The amino acid sequence appearing between these consensus regions is precisely that, in every case, set forth in Figure 3 as the appropriate intrapolypeptide linker. (An extra copy of Figure 3 is included as Exhibit B to facilitate comparison.)

Figure 3 of the patent shows the N-terminal portions of the interpeptide linkers that are at the starts of module 3 and module 5. Unfortunately, the Donadio article does not show the complete upstream sequence from the KS domain at the start of these proteins as indicated by the position numbers at the right of the figure; there is space for 100 amino acids in each line and the numbers at the ends of these lines are 126 and 122, respectively. However, it will be seen that the sequence immediately preceding the KS domain, in each case, matches that presented in the figure. The additional upstream sequence is shown in Exhibit C-1; the complete upstream sequence known for erythromycin matches that described herein. The nature of the sequence could, of course, be obtained by reference, for example, to PCT publication WO 93/13663 which would provide this information, or to the GenBank deposit referred to in Donadio. Similarly, Figure 3 does not provide information on the C-terminal portion of the interpolypeptide linkers, but this too could be obtained from the PCT publication and is shown in Exhibit C-2. Again, the figure in the Donadio paper is not quite complete - for example, DEBS-1 that would contain the

C-terminal portion downstream of the ACP of module 2 contains 3491 amino acids while the depiction in Donadio shows only up to position 3418. The remainder of this sequence can also be derived from the GenBank deposit referenced in the Donadio paper, for example.

Exhibit D shows additional RAL sequences for rapamycin and erythromycin; the N-terminus of the KS domain is readily determined in all of these PKS proteins as an acidic amino acid (E or D) followed by proline, followed by a hydrophobic amino acid (L, I or V). The linker sequences in the rapamycin stack are much more homologous than the erythromycin linkers. By comparison of the termini of the KS domain, however, between the erythromycin and rapamycin sequences, the linker sequences for rapamycin can be determined. With respect to the intramolecular sequences, M3*rap* is that designated in the Figure ACP02/KS3; that designated M4*rap* is that designated as ACP03/KS4*rap*; that designated M7*rap* is that set forth as ACP06/KS7. (Exhibit D does not depict any REL sequence.)

With respect to the comments regarding GenBank Accession No. M63676, the approximate values of module 1 and module 2 should be taken as simply "approximate." As shown by the consensus sequence in the Donadio article, which depicts the sequence of M63676, the amino acid sequences presented in Figure 3 are correct.

In view of the foregoing illustration, it is believed that this basis for rejection may be withdrawn.

The Rejection Under 35 U.S.C. § 112, Paragraph 1

All claims were rejected on this basis for an asserted lack of written description. The Examiner is, of course, correct that if a composition of matter is claimed, there must be an adequate description of the components that are required. Applicants respectfully submit, however, that in view of the showing set forth above with regard to the rejection under § 112,

paragraph 2, this is indeed the case. By simply consulting the known sequences of any cloned PKS, the relevant structure of the linkers can be ascertained in precisely the same manner as can the naturally occurring catalytic domains of the modules themselves.

The attention of the Office is drawn, for example, to the holding in *Amgen, Inc. v. Hoechst, Marion Roussel, Inc.*, 57 USPQ2d 1449 (DC Mass. 2001), 65 USPQ2d 1385 (Fed. Cir. 2003). As set forth in that case, the written description requirement does not necessarily require spelling out sequences of amino acids or nucleotides, provided the structural features can be ascertained in a sufficient manner to permit one of ordinary skill in the art to construct them. This is clearly the case here since one has only to review published sequences of modular PKS to define the linkers that are present therein. The fact that these linkers happen to be disparate in structure from each other is an accident of nature and does not preclude the ordinary skilled artisan from ascertaining workable embodiments with ease.

In view of the foregoing, it is believed that the written description requirement is met.

CONCLUSION

As the lack of consensus among intermolecular and intramolecular linkers is an accident of nature and not the fault of applicants, and as it would be well within ordinary skill to obtain workable linkers for construction of the claimed compositions, applicants believe that the pending claims, claims 23, 25-26 and 28-39 are in a position for allowance and passage of these claims to issue is respectfully requested.

In the unlikely event that the transmittal letter is separated from this document and the Patent Office determines that an extension and/or other relief is required, applicants petition for any required relief including extensions of time and authorize the Assistant Commissioner to

charge the cost of such petitions and/or other fees due in connection with the filing of this document to **Deposit Account No. 03-1952** referencing docket No. 300622004600.

Respectfully submitted,

Dated: May 5, 2003

By: Kate H. Murashige

Kate H. Murashige
Registration No. 29,959

Morrison & Foerster LLP
3811 Valley Centre Drive,
Suite 500
San Diego, California 92130-2332
Telephone: (858) 720-5112
Facsimile: (858) 720-5125

EXHIBIT A. - VERSION WITH MARKINGS TO SHOW CHANGES MADE

In the Claims:

26. (Twice amended) The hybrid modular PKS of claim 23 wherein the portion of the ERL at the N-terminus of the second module is selected from the group consisting of [M3 *ery*, M5 *ery*, M4 *rif*, M7 *rif*, M8 *rif*, M9 *rif*, M5 *rap*, and M11 *rap* inter-module linkers wherein the portions of said modules coupled to the N-terminus of the succeeding module are represented by] SEQ. ID. NO's: 12-19, respectively.

28. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains *ery* modules 1 and 3 through 6 inclusive and tylosin module 2, and wherein said polyketide chain is transferred from *ery* module 1 to *tyl* module 2 and then to *ery* modules 3 through 6 inclusive.

29. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains *ery* modules 1 through 5 inclusive and narbomycin module 6, wherein said polyketide chain is transferred from *ery* modules 1 through 5 inclusive to *nar* module 6.

30. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains modules 1 and 3 through 6 inclusive of *ery* and modules 2-3 of tylosin, spiramycin or niddamycin, wherein said polyketide chain is transferred from *ery* module 1 to modules 2-3 of tylosin, spiramycin or niddamycin and then to *ery* modules 3 through 6 inclusive.

31. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains modules 1 through 3 inclusive of tylosin, spiramycin or niddamycin and modules 3 through 6 inclusive of *ery*, and wherein said polyketide chain is transferred from modules 1 through 3 inclusive of said tylosin, spiramycin or niddamycin to *ery* modules 3 through 6 inclusive.

32. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains a module of tylosin, spiramycin or niddamycin and modules 1-2 and 3 through 6

inclusive of *ery*, wherein said polyketide chain is transferred from *ery* modules 1-2 to the tylosin, spiramycin or niddamycin module and then to *ery* modules 3 through 6 inclusive.

33. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains modules 1 and 3 through 6 inclusive of *ery* and module 5 of tylosin, spiramycin or niddamycin having the enoyl reductase catalytic activity inactivated, wherein said polyketide chain is transferred from *ery* module 1 to module 5 of tylosin, spiramycin or niddamycin and then to *ery* modules 3 through 6 inclusive.

34. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains *ery* modules 1 through 4 inclusive and 6 and module 6 of spiramycin or niddamycin, wherein said polyketide chain is transferred from *ery* modules 1 through 4 inclusive to module 6 of spiramycin or niddamycin and then to *ery* module 6.

35. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains module 1 of FK-506 or 520 and modules 2 through 14 inclusive of rapamycin, wherein said polyketide chain is transferred from module 1 of FK-506 or 520 and then to modules 2 through 14 inclusive of rapamycin.

36. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains module 1 and 11 through 14 inclusive of rapamycin and modules 2 through 6 inclusive of FK-506 or 520 wherein said polyketide chain is transferred from module 1 of rapamycin to modules 2 through 6 inclusive of FK-506 or 520 and then to modules 11 through 14 inclusive of rapamycin.

37. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains module 1 of rapamycin, modules 2 through 7 inclusive of FK-506 or 520 and modules 12 through 14 inclusive of rapamycin, wherein said polyketide chain is transferred from module 1 of rapamycin to modules 2 through 7 inclusive of FK-506 or 520 and then to modules 12 through 14 inclusive of rapamycin.

38. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains module 1 of rapamycin, modules 2 through 8 inclusive of FK-506 or 520 and modules 13-14 of rapamycin, wherein said polyketide chain is transferred from module 1 of rapamycin to modules 2 through 8 inclusive of FK-506 or 520 and then to modules 13-14 of rapamycin.

39. (Twice amended) The hybrid modular [polyketide] PKS of claim 23 which contains modules 1 through 10 inclusive of rapamycin and modules 7 through 10 inclusive of FK-506 or 520, wherein said polyketide chain is transferred from modules 1 through 10 inclusive of rapamycin to modules 7 through 10 inclusive of FK-506 or 520.

Gene. 111 (1992) 51–60

GENE 06278

Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*

(FAS; fatty acids; macrolide antibiotic; sequence alignments; *Streptomyces*)

Stefano Donadio and Leonard Katz

Corporate Molecular Biology, Abbott Laboratories, Abbott Park, IL 60064 (U.S.A.)

Received by C.R. Hutchinson: 21 August 1991
Revised/Accepted: 15 October/17 October 1991
Received at publishers: 12 November 1991

SUMMARY

Localization of the enzymatic domains in the three multifunctional polypeptides from *Saccharopolyspora erythraea* involved in the formation of the polyketide portion of the macrolide antibiotic erythromycin was determined by computer-assisted analysis. Comparison of the six synthase units (SU) from the *eryA* genes with each other and with mono- and multifunctional fatty acid and polyketide synthases established the extent of each β-ketoacyl acyl-carrier protein (ACP) synthase, acyltransferase, β-ketoreductase, ACP, and thioesterase domain. The extent of the enoyl reductase (ER) domain was established by detecting similarity to other sequences in the database. A segment containing the putative dehydratase (DH) domain in EryAII, with a potential active-site histidine residue, was also found. The finding of conservation of a portion of the DH-ER interdomain region in the other five SU, which lack these two functions, suggests a possible evolutionary path for the generation of the six SU.

INTRODUCTION

Erythromycin, a macrolide antibiotic produced by *S. erythraea*, is composed of the polyketide-derived 14-membered macrolactone ring, 6dEB, to which are attached two deoxysugars, cladinose and desosamine. Synthesis of 6dEB involves six elongation cycles that resemble the steps in fatty acid synthesis. It has recently been shown that 6dEB synthesis requires three adjacent *eryA* genes encoding large multifunctional polypeptides and that the *eryA* cluster consists of six modules (repeated motifs), each encoding a different SU specific for one of the elongation steps (Cortes et al., 1990; Donadio et al., 1991). We proposed that the genetic organization of *eryA* and the steps in the biochemical pathway of 6dEB are colinear (Donadio et al., 1991). A scheme of the enzymatic activities leading to 6dEB is shown in Fig. 1.

In previous work, the FAS-like activities ACP, AT, KR and KS in the multifunctional *eryA*-encoded polypeptides were identified from the presence of 'signature sequences' found at the active site of the functional domains (Cortes et al., 1990; Donadio et al., 1991). However, signature sequences have not been assigned previously to the DH and ER functions, both of which have been poorly characterized biochemically. Multifunctional FAS systems are or-

Correspondence to: Dr. L. Katz, Abbott Laboratories, D-93D, One Abbott Park Rd., Abbott Park, IL 60064 (U.S.A.)
Tel. (708)937-4132; Fax (708)938-6046.

Abbreviations: aa, amino acid(s); ACP, acyl-carrier protein; ACP-S, ACP of SU1; AT, acyltransferase; CoA, coenzyme A; 6dEB, 6-deoxyerythronolide B; DH, dehydratase; *dnaB*, gene encoding helicase; ER, enoyl reductase; *ery*, erythromycin biosynthesis gene; *eryA*, gene encoding 6dEB synthase; FAS, fatty acid synthase; FAS1, *S. cerevisiae* FAS β-chain; KR, β-ketoreductase; KS, β-ketoacyl ACP synthase; 6MSAS, 6-methylsalicylic acid synthase; ORF, open reading frame; PKS, polyketide synthase; *S.*, *Saccharopolyspora*; SU, synthase unit(s); TE, thioesterase; URF, unidentified ORF.

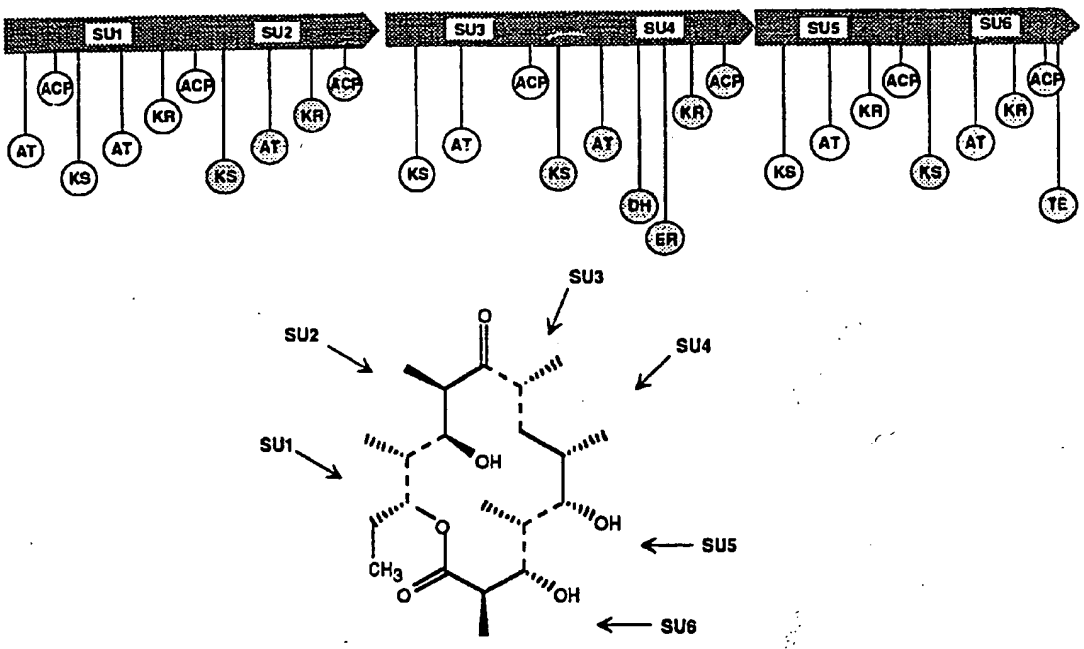


Fig. 1. Scheme of 6-deoxyerythronolide B synthesis (Donadio et al., 1991). The top portion shows the three *eryA*-encoded polypeptides containing SU 1 through 6. Enzymatic activities belonging to the first and to the second SU for each polypeptide are represented by empty and shaded circles, respectively. The bottom portion illustrates the role of each SU in the synthesis of 6dEB, where the C_2 units in the ring introduced by odd- and even-numbered SU are represented by dashed and continuous lines, respectively.

ganized in discrete functional domains which can be resolved upon limited proteolysis (Wakil, 1989). The *eryA*-determined SU can be assumed to consist of linear domains on the basis of their similar organization to the animal FAS systems (Cortes et al., 1990; Donadio et al., 1991). Here, by comparing the 6dEB PKS domains with each other and with other multi- and monofunctional FAS and PKS systems, we subdivide each *eryA*-encoded polypeptide into its constituent domains and propose a location for the DH and ER domains in EryAII. Conclusions similar to those reported here have been independently reached in the laboratory of P.F. Leadlay (personal communication) for the domain organization of EryAII and EryAIII and by Witkowski et al. (1991a) for the domain organization of rat FAS.

RESULTS AND DISCUSSION

(a) Extent of KS, AT and ACP domains

The six SUs are organized in pairs in three large deduced aa sequences (Fig. 1). Within each polypeptide, the end of the N-terminal SU was arbitrarily placed where the sequence C-terminal to the first ACP domain began to diverge from that of the second ACP, located toward the C-terminal end of the polypeptide. In this way, SU 1 through 6 were determined to be 1975, 1516, 1484, 1973, 1480 and 1690 aa in length, respectively. The domain order f each SU is KS, AT, KR and ACP. SU1 has additional AT and ACP domains at its N terminus, SU4 contains DH and ER domains between AT and KR, and SU6 contains a TE at its C-terminal end (Fig. 1). We have used the different compositions of the SU as a first approximation in establishing the extent of some of the domains and confirmed the results obtained by comparison with the monofunctional PKS proteins from the *gra* (Sherman et al., 1989) and *tcm* (Bibb et al., 1989) clusters. In this way, the beginnings and ends of the KS domains could be easily assigned to the highly conserved aa motifs d(e)PiAiVgmaCR (uppercase letters refer to invariant residues) and GTNAHvIeE, respectively (Fig. 2). Comparison of the sequence of the first AT of module 1 (AT-S) to the previously aligned six other ATs clearly indicated the aa motif vfvFPGQGaQW as the likely beginning of the AT domains (Fig. 2). In a similar way, their end could be placed where AT-S is seen to diverge from the other ATs, a few residues after the highly conserved aa motif GVavdwxxa (Fig. 2). Comparison of the *eryA* ATs with the only monofunctional sequence available to us, a transacylase from *Streptomyces glaucescens* (R. Summers and C.R. Hutchinson, personal communication), confirmed the assignment of the N-terminal end and showed significant matches, albeit with two long gaps, up to the pvxLPt motif, just beyond the C-terminal end of the domains established by comparison of the *eryA* ATs.

When the segment encompassing the ACP active-site aa motif LGxDS from the first ACP of SU1 (ACP-S) was aligned with the six other *eryA* ACPs, the ends of the ACP domains appeared to coincide with the aa m tif lAxhlxa,

situated 38 aa after the active-site Ser (Fig. 2). Their starts could not be easily established by this criterion, however, since conservation between ACP-S and the other six ACPs began a few aa before the active-site motif. Comparison with monofunctional ACPs from *gra* and *tcm*, however, indicated that significant matches began with the motif lAglsxxe and ended with the motif Glrlpxtlv, 15 aa before the end established by comparison of the *eryA* ACPs alone (Fig. 2). This would place the N-terminal end in all of the ACP domains, except ACP-S, 45 aa upstream from the active-site Ser.

(b) Conservation of KS, AT and ACP in multifunctional systems

The aligned KS, AT and ACP domains from the six *eryA* SUs were compared with those found in other multifunctional systems, FAS from chicken (Yuan et al., 1988; Holzer et al., 1989) and rat (Amy et al., 1989), and the PKS 6MSAS from *Penicillium patulum* (Beck et al., 1990). As expected, a higher overall degree of similarity was observed in the intra- than in the interdomain regions among the multifunctional systems analyzed (Fig. 2). The nine KS domains examined shared 71 invariant aa residues (out of 425 aa) and, considering conservative substitutions, were similar for 205/425 residues (Fig. 2). In contrast, AT and ACP domains exhibit lower conservation (Fig. 2; see also below). It is tempting to speculate that the basis for the apparently high sequence constraint in the KSs from distantly related organisms is that, in addition to catalyzing the condensation of the acyl chain with the extender unit charged on the ACP to form β-ketoacyl-ACP (Wakil, 1989), the KS is also responsible for the transacylation of the elongated acyl chain from the ACP to its own active-site Cys residue.

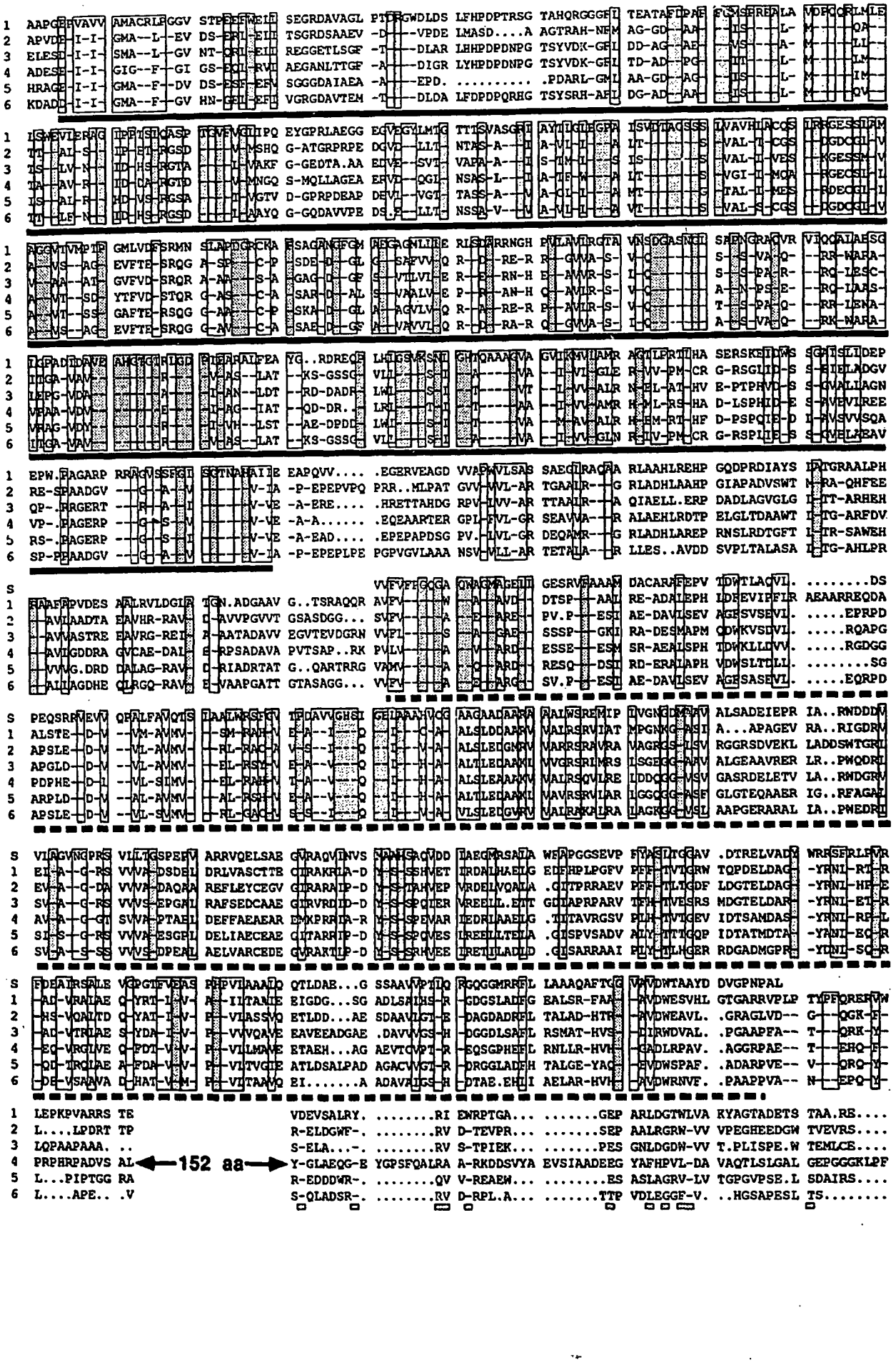
Invariant aa accounted for 26/345 residues in the ten AT domains examined, and approx. 30% of the ten sequences involved conservative substitutions (Fig. 2). It is noteworthy that, in addition to the segment around the signature aa sequence GHSxG, two additional segments, each containing an invariant His, are highly conserved in the ten ATs. Serine proteases are known to contain active-site Ser and His residues, distant in the primary structure, but brought into close proximity in the folded protein (Hess, 1971). Since the types of reaction carried out by ATs and serine proteases are believed to be similar (McCarthy and Hardie, 1984), the finding of invariant His in the ATs distant from the active-site Ser suggests a similarity with serine proteases also in catalytic mechanism.

The ten ACP domains examined exhibit only one invariant residue outside of the LGxDS motif. The 30-aa N-terminal segment of the ACP domains from the SU shows some conservation with mono- and multifunctional proteins, except for ACP-S (data not shown). In their independent study of domain organization of rat FAS, Witkowski et al. (1991a) have placed the N-terminal end of the ACP domain approx. 10 aa C-terminal to the end suggested here. Thus, the ACP-S domain lacks in its N-terminal portion a segment of at least 15 aa when compared to the other ACPs. This apparent anomaly may reflect a functional difference between ACP-S and the other ACPs. According to the model proposed for 6dEB synthesis (Donadio et al., 1991), the sole role of ACP-S consists of receiving the propionyl starter unit from AT-S and of transferring it to the KS of SU1. Its function would thus be limited to acylation/deacylation, and this ACP would not be employed in carrying the β-ketoacyl chain through the appropriate processing steps, as do all other FAS and PKS ACPs known.

(c) Extent of ER, DH, KR and TE domains

Only *eryA* module 4 encodes DH and ER functions, which, to date, have only been tentatively located in FAS systems (Tsukamoto and Wakil, 1988). A 400-aa segment unique to SU4 and believed to include the ER domain (Donadio et al., 1991), was used to search the databases. Surprisingly, the best matching sequences found, aside from the rat and chicken FASs, were from structural proteins of higher eukaryotes, ζ-crystallin from guinea pig lens (Rodokanaki et al., 1989) and the membrane protein VAT-1 from *Torpedo californica* synaptic vesicles (Linial et al., 1989). The similarity of these two proteins to alcohol dehydrogenases has already been reported. In addition, the 5′ end of an URF divergently transcribed from the *dnaB* gene of *Salmonella typhimurium* (Wong et al., 1988) was detected by this search. Alignment of these sequences indicated that the ER domain is likely to extend for approx. 330 aa and contains 19 invariant and 90 conserved aa residues (Fig. 3). In particular, the sequence LxHxg(a)xGGVG, proposed as the NADPH-binding site for the rat ER (Amy et al., 1989; Witkowski et al., 1991a), appears to be highly conserved in the six sequences examined. It should be noted that database searches also indicated similarity between the ER domain and alcohol dehydrogenases, but this similarity is limited mainly to the putative NADPH binding site (data not shown). Although no enzymatic role has been assigned to ζ-crystallin or VAT-1, the high similarity detected among the six sequences suggests a possible present or former role for the two monofunctional proteins in reducing double bonds that lie α,β to a carbonyl group (Piatigorsky and Wistow, 1991).

The ca. 500-aa segment defined by the end of the AT and the beginning of the ER domain in SU4 showed some similarity to the corresponding segments from rat and chicken FAS (Donadio et al., 1991), as well as with an approx. 200-aa stretch located between the AT and KR domains of 6msas (data not shown), which is believed to


152 aa

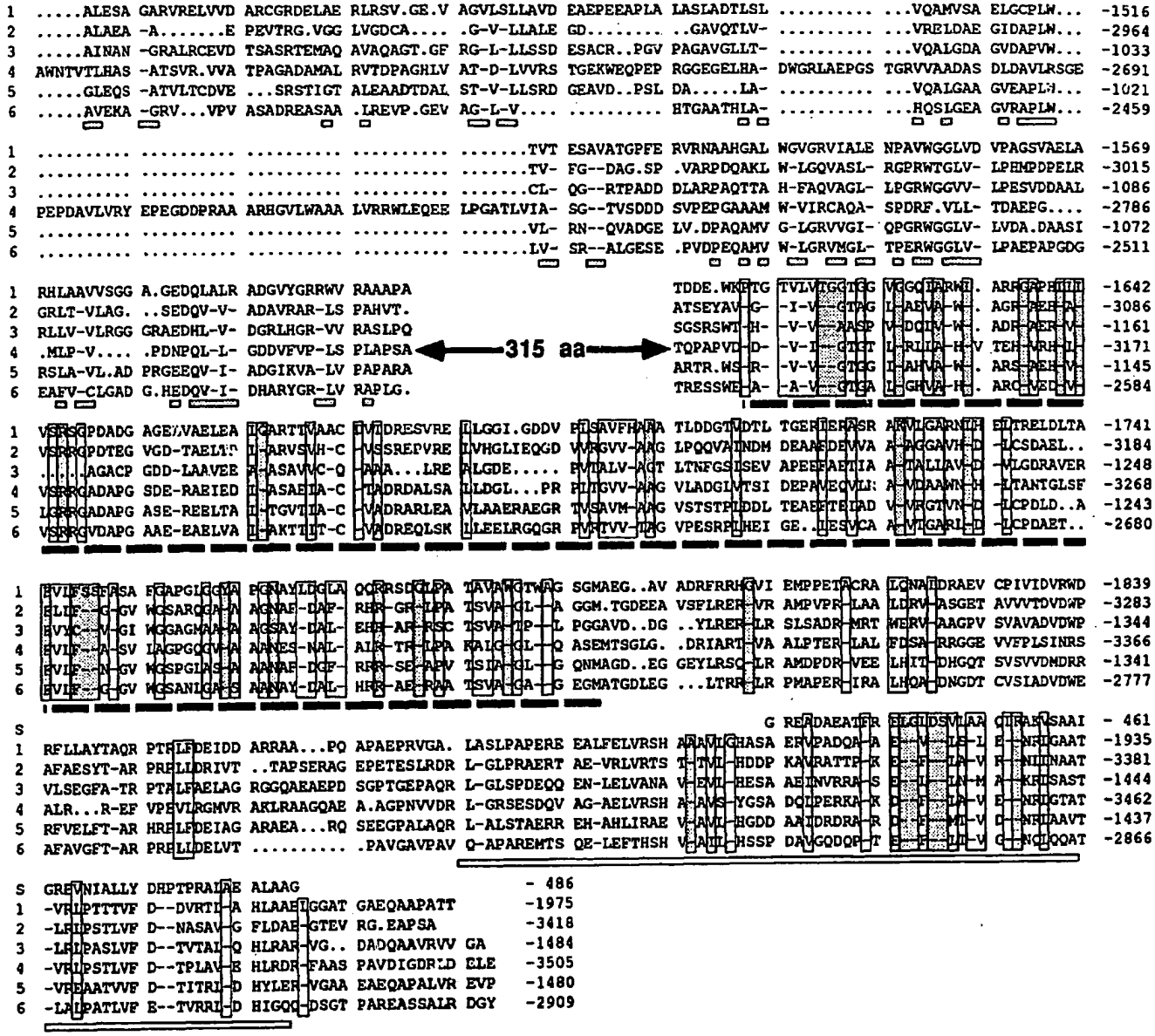


Fig. 2. Alignments of the six *eryA* SU. The symbols on the left margin refer to the particular SU, with S referring to AT-S or ACP-S, as appropriate. Numbers on the right margin refer to the aa sequence position at the end of each row in EryAI (for 1, 2 and S on left), in EryAII (for 3 and 4) and in EryAIII (for 5 and 6). Sequences for EryAI and for EryAII and -III are from GenBank, accession Nos. M63676 and M63677, respectively. Invariant aa residues in the six SU are marked by dashes. Dots refer to computer-introduced gaps to maximize alignments. Shaded boxes refer to aa residues invariant in the six (or seven) sequences from the SU, as well as chicken FAS (Holzer et al., 1989; Yuan et al., 1988), rat FAS (Amy et al., 1989) and 6MSAS (Beck et al., 1990). Open boxes refer to conservative substitutions or invariant residues in all but one sequence. The N terminus of chicken FAS is assumed to precede the published sequence (Holzer et al., 1989), as recently reported (Witkowski et al., 1991a). The KR of SU3, when it deviates from the other eight sequences, is ignored for boxing purposes. The extent of each domain is indicated by underlining of the sequences with solid black bars, short, heavy dashes, long, heavy dashes, and open bars, representing the KS, AT, KR and ACP domains, respectively. The two arrows mark the extra segments of 152 and 315 aa present in SU4, which are presented in Figs. 4 and 3, respectively. The shaded bars under the sequences in the region comprised between the two arrows indicate invariant and conservative substitutions among the six SU. Computer-assisted sequence analyses were performed using the University of Wisconsin GCG programs (Devereux et al., 1984). Sequences were examined pairwise using COMPARE/DOTPLOT. Multiple sequence alignments were performed using PILEUP, with a gap weight of 3.0 and a gap length weight of 0.1. The sequences of the six SU were initially aligned. Subsequently, the segments corresponding to the first AT and ACP of SU1 (AT-S and ACP-S, respectively) were individually aligned with the other six AT and ACP domains, respectively. Finally, the region of SU4 between the DH and ER domains (section e) and those from SU 1, 2, 3, 5 and 6 between the AT and KR domains were separately aligned. The three alignments generated in this way were manually combined using LINEUP. For comparing the six *eryA* SU with the other multifunctional systems examined, chicken FAS, rat FAS and 6MSAS, PILEUP was run using one sequence from each group (one SU, one FAS and 6msas) or with all nine sequences, with similar results. The DH-ER interdomain region of SU4 and the AT-KR interdomain regions from the other five SU, when compared with the other multifunctional systems, gave substantially different alignments upon changing of the PILEUP parameters. These segments have thus been ignored for boxing purposes.

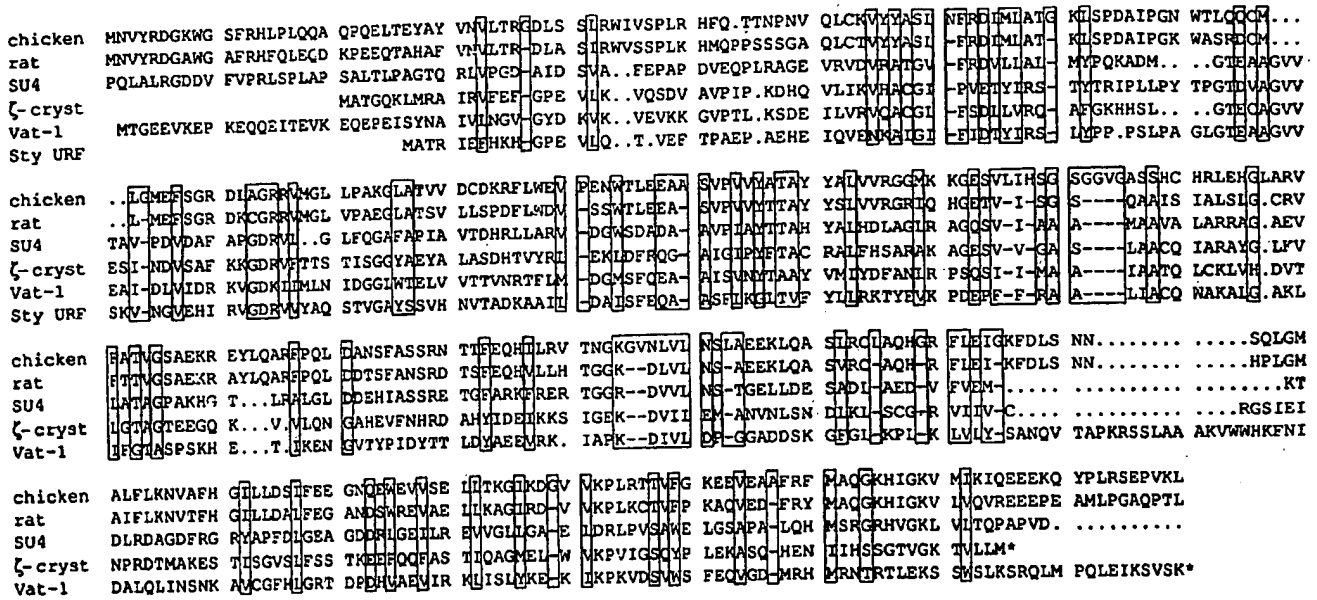


Fig. 3. Putative enoyl reductase domains. The segment from EryAII comprised between aa 2832 and 3138 was employed to screen the GenBank and EMBL (48285 sequences) and the Swissprot (20722 sequences) databases using the programs TFASTA and FASTA, respectively. Sequences showing significant matches were aligned using PILEUP. Invariant aa residues are represented by dashes; dots refer to inserted gaps. Conserved regions among all the sequences are boxed. An asterisk indicates the stop codon in the corresponding gene. Sequences as indicated: chicken, FAS from chicken, aa 1483–1853; rat, FAS from rat, aa 1496–1866; SU4, *S. erythraea* EryAII, aa 2795–3138; ζ-cryst, guinea pig lens crystalline, complete sequence (Rodokanaki et al., 1989); VAT-1, membrane protein from *Torpedo californica* cholinergic synaptic vesicles, complete sequence (Linial et al., 1989); Sty URF, *Salmonella typhimurium* unidentified open reading frame divergent from *dnaB*, translated from GenBank J03390 (Wong et al., 1988). Note that only the 5′ end of this sequence is available, with the resulting polypeptide ending as KALGAKL^{168}.

contain a DH function (Beck et al., 1990). The alignment of these four sequences (Fig. 4) indicates that significant homology is limited to an approx. 150-aa segment. Within it, the invariant HxxxGxxxxP motif is embedded in a 25-aa segment with a high degree of conservative substitutions, involving mostly hydrophobic residues. The finding of an invariant His residue in the most conserved region among the four sequences is consistent with the proposed role for a His as the active-site residue in the *E. coli* β-hydroxydecanoyl thioester dehydrase (Bloch, 1971). The corresponding gene encodes the active-site His in the sequence HFIGDPVMP^{78} (Cronan et al., 1988), where insertion of a gap between I and G would conform this sequence to the proposed consensus. The HxxxGxxxxP motif is also found at a single position in the *S. cerevisiae* FAS1 sequence as HLSHGVKMIP^{1057}, approximately where the DH domain has been tentatively placed (Schweizer et al., 1986; Chirala et al., 1987). These observations suggest that the DH domain in multifunctional FAS and PKS systems is relatively short, extending for approx. 140–170 aa, consistent with the 170-aa size of the *E. coli* enzyme, and point to a specific His as one of the active-site residues involved in catalysis. The same His has been independently proposed by P.F. Leadlay (personal communication) as the catalytic residue in the DH domain of EryAII. It should also be noted that the two animal FASs and SU4 also share the motif GYxYGPxFQ, approx. 110 aa after the proposed active-site His, whereas 6msas does not. The

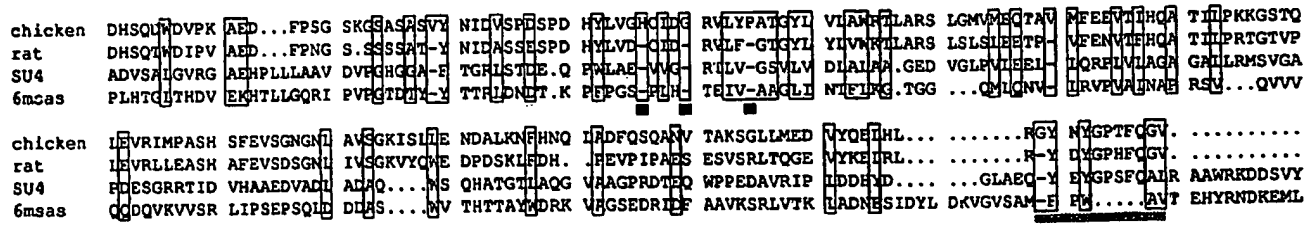


Fig. 4. Putative dehydratase domains. The approx. 500-aa segments from EryAII, chicken FAS, rat FAS and 6msas suspected to contain the DH domain (see section c) were aligned using PILEUP. Only the portion showing significant matches is represented. For abbreviations and symbols, refer to Fig. 3. The three blackened squares denote the putative active-site motif HxxxGxxxxP. The shaded bar denotes the highly conserved region common only to the two FAS sequences and SU4. Sequences: chicken, FAS from chicken, aa 812–987; rat, FAS from rat, aa 837–1009; SU4, *S. erythraea* EryAII, aa 2365–2551; 6msas, *Penicillium patulum* 6MSAS, aa 914–1096.

reason for this difference and a possible role for this motif are at present unknown. No significant matches were detected by database searching with the proposed DH domain or by comparison with known aa dehydratases.

All six *eryA* SU contain a segment corresponding to a KR domain, although the KR domain of SU3 is believed to be non-functional (Donadio et al., 1991). The beginning of the KR domains is likely to coincide with the region following the ER domain where SU4 realigns with the other five SU. This location was matched in the N-terminal portion of the monofunctional KRs involved in actinorhodin (Hallam et al., 1988) and granaticin (Sherman et al., 1989) synthesis. Thus, the *ery* KR domains are likely to start with the PxGTvLv motif, just upstream from the putative NADPH-binding site (Fig. 2). Since in all multifunctional FAS and PKS systems the KR is always followed by an ACP, the end of the KR domains was placed approx. 190 aa after the NADPH-binding site, where conservation among the nine sequences examined began to decline (Fig. 2). This interpretation results in the separation of the KR and ACP domains of 90–100 aa in the PKS systems and of 60 aa in the FAS systems.

The C-terminal end of SU6 contains a TE domain. This domain was compared with the corresponding domains from the two FAS sequences, with monofunctional thioesterases from rat (Randhawa and Smith, 1987; Safford et al., 1987) and duck (Poulose et al., 1985), and with the TE-like ORF downstream from *eryF* (Weber et al., 1991). The alignments (Fig. 5) indicate that the TE domain in *eryA* extends for approx. 230 aa, and includes, in addition to the invariant GxSxG motif common to ATs and serine proteases, the GdH motif found near the C-terminal end, which has been shown by site-directed mutagenesis to be essential for activity (Witkowski et al., 1991b). Overall, little similarity was detected among the six TEs analyzed, which may be related to the three different classes of substrates recognized by these enzymes (Wakil, 1989; Cortes et al., 1990; Donadio et al., 1991). This is exemplified by the low similarity between the TE domain of rat FAS and the short chain TE from the same organism, and between the TE of SU6 and the other TE-like sequence from the *ery* cluster, although the role of the latter has not yet been determined.

(d) Inter- and extradomain regions

The overall domain organization of the three *eryA*-encoded polypeptides is summarized in Fig. 6. It can be seen that the largest interdomain regions are the five segments between AT and KR, and the one between DH and ER in EryAII. When the AT-KR interdomain regions from SU 1, 2, 3, 5 and 6 were compared with the region from SU4 containing the DH and ER domains, some similarity could be detected under relatively stringent conditions (data not shown). Computer-generated alignment of these six segments indicated that these regions can be best accommodated after accounting for two insertions in SU4, the first of 152 aa, and the second of 315 aa (Fig. 2). These two insertions correspond very closely to the DH and ER domains, respectively, as determined above. In the 200-aa segment which joins the DH and ER domains in SU4, a

```
rat       AQLNLSILLV NPEGPTLTRL NSVQSSERPL FLVHPIEGSI TV...FHSLA AMLS..... .MPTYGLQCT QAAFL.DSIP NLAAYYIDCI KQVQPEGPHR
chicken   PKLDLNNLLV NPEGPTITRL NEVQSTERPL FLVHPIEGSI AV...FYTLA SKIH..... .MPCYGLQCT KAAFL.DSIQ SLASYYIDCM KQIQPEGPYR
rat SC    METAVNAKSP RNEKVLNCLY QNPDAVFKLI CFPWAGGGSI H....FAKWC QKINDSLEVH AVRLAYPTR LGEPFANDIY QIADEIVTAL LPIIQDKPFA
duck SC               MDKVIARPY KRPNALCRLI CFPWAGGNCS F....FIPWC EAFSSIIVVS VIRLAYPECR DTEPFPEDMA EVVNEITNAL LKDLQEKPFA
ery ORF               MSTWLRRFG PPVEHRARLV CFPHAGAAAD S....YLDLA RALAPEIDVH AVQYPGRQDR RDEEPLGTAG EIADEVAAVL RASGGDGPFA
SU6       GLSDFREHFD GSDGFSLDLV DMADGPGEVT VICQAGTAAI SGPHEGTFLA GALRGIAPVR ALPQPGYE.. EGFGLPSSMA AVAAVQADAV IRTQGDKGFV

rat       VAGYSFGACV AFEMCSQLQA QQGPAPAHNN LFLFDGSHTY VLAYT..... ...QSYRAKL TPGCEAEAEA EALCFFIKQF VDAEHSKVLE ALPLKSLED
chicken   IA-Y-F-ACV AFEMCSQ-QA QQNASHALNS LFLFDGSHSF VAAYTQCFSF SLFQSYRAKL TQGNEAALET EALCAFVQQF TGIEYNKVLE IILPLEDLEA
rat SC    FF-H-F-SYI ALITALL-KE KYKMEPLH.. IFVSGASAPH STSRPQVPDL ..........  .....NELTE EQVRHHLLDF GGTPKHLIED QDVLRMFIPL
duck SC   LF-H-F-SFM SYALAVH-KE KHGLEPVH.. MFFSGSYGPH SEYFHLMYKL ..........  .....PEVED SRILELIHTL GGTPPEFLQN EQTTKHLLRV
ery ORF   LF-H-M-ALI AYETARR-ER EPGGGPLR.. LFVSGQTAPR VHERR..TDL ..........  .....PG..D DGLVDELRRL .GTSEAALAD EALLAMSLPV
SU6       VA-H-A-ALM AYALATE-LD RGHPPRGV.. VLLDV..... .........Y ..........  .....PPGHQ DAANAWLEEL TATL...LEDR ETV.....RM

rat       RVAAAVDLIT RSHQSLDRRD LSFAAVSFYY KLRAADQYKP KAKYHGNVIL LRAKTGGTYG EDLGADYNLS QVCDGKVSVH IIEGDHRTLL EGRGLESIIN
chicken   RVNAAADLIT QIHKNINREA LSFAAASFYH KLKAADKYIP ESKYHGNVTL MRAKTHNEYE EGLGGDYRLS EVCDGKVSVH IIE-D-RTLL EGDGVESIIG
rat SC    LKADAGVVKK FIFDKPSKAL LSLDTTGFL. .......... .......... ....GSEDTI K...DIEGMQ DLTSGKFDVH MLP-D-FYLM KPDNENFIKN
duck SC   LKEDQKVIVT YPWHDVRKKY FSCDITCFN. .......... .......... ....GSDEKN H...GSEAWI AITSGDTSIY SLP-N-FYLM EPSNETFLIK
ery ORF   LRADYRVLRS YAWADGPP.. LRAGITALC. .......... .......... ....GDADPL TATGDAERWL QHSVIPGRTR TFP-G-FYLG EQVTE..VAG
SU6       DDTRLTALGA YDRLTGQWRP RETGIFTLL. .......... .......... ....VSAGEP MGPWPDDSWK PTWPFEHDTV AVP-D-FTMV QEHADAIARH

rat       IIHSSLAEPR VSVREG*
chicken   IIHGSLAEPR VSVREG*
rat SC    MIAKCLELSS LT*
duck SC   MITKCIENSD I*
ery ORF   AVRRDLLRAG LAG*
SU6       IDANLGGGNS *
```

Fig. 5. Thioesterase domains. The six TEs were compared using PILEUP. See legend to Fig. 3 for symbols. Sequences: rat, FAS from rat, aa 2209–2505; chicken, FAS from chicken, aa 2193–2497; rat SC, short-chain TE from rat, complete sequence (Randhawa and Smith, 1987; Safford et al., 1987); duck SC, short-chain TE from duck, complete sequence (Poulose et al., 1985); eryORF, downstream from *S. erythraea eryF* (Weber et al., 1991; GenBank accession No. M54983); SU6, *S. erythraea* EryAIII, aa 2927–3170.

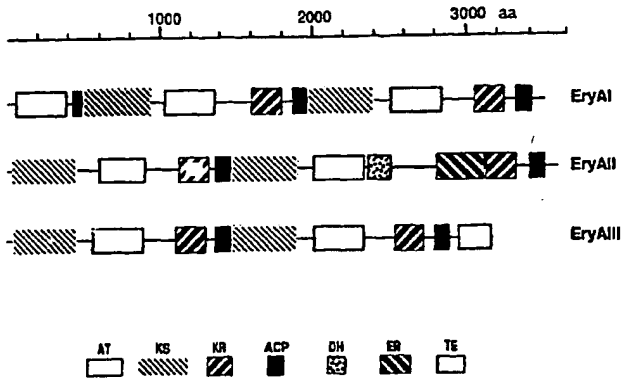


Fig. 6. Domain organization in 6dEB synthase. Each domain is represented by a rectangle of different filling as shown, whose length is proportional to the length of the domain. Note the partial filling of the first KR of EryAII, which denotes an inactive KR (Donadio et al., 1991).

stretch of approx. 80 aa appears to be fairly well conserved among the six SU (Fig. 2). The enzyme from *S. erythraea* AKR5, which has been deleted of this 80-aa segment along with the KR domain C-terminal to it in SU5, retains active SU5 and SU6, as judged by the ability to produce a significant amount of the 6dEB analog lacking the hydroxyl group introduced during synthesis step 5 (Donadio et al., 1991; Fig. 1). Whilst Witkowski et al. (1991a) have speculated that the long DH-ER interdomain segment is involved in facilitating protein–protein interactions in the dimeric FAS enzyme, our results indicate that the presence of at least a portion of this segment is not absolutely required for 6dEB PKS function, although the level of activity of the altered enzyme could not be measured directly. That the deleted segment is important for KR activity is improbable, since very little homology was detected with other multifunctional systems in this region (data not shown).

An additional feature of the *eryA*-encoded polypeptides is the presence of extra N-terminal and C-terminal tails extending significantly beyond the domain limits (Fig. 6). The N termini of polypeptides EryAII and EryAIII contain segments of 26 and 33 aa preceding the KS domains of SU3 and SU5, respectively, and the C termini of EryAI and EryAII contain stretches of 69 and 63 aa following the ACP domains of SU2 and SU4, respectively. The other multifunctional systems examined do not contain extra tails of such lengths. In 6msas a 28-aa segment precedes the KS domain, but the ACP domain is followed by a segment of only 6 aa at the C terminus of the polypeptide. The KS domain of rat FAS starts at the N terminus of the protein. It is tempting to speculate that these additional segments in the *eryA*-encoded polypeptides may play a role in facilitating the correct intermolecular transfer of the growing acyl chain, such as from SU2 in EryAI to SU3 in EryAIII, either by enabling specific protein–protein interactions, or by properly positioning the polypeptides on some cellular structure.

(e) Ev lution of the modules

S. erythraea contains a Type-I PKS and, most likely, a Type-II FAS system (Revill and Leadlay, 1991). The evolutionary origin of these two systems can be understood by comparison of similar enzymatic functions belonging to a Type-I or Type-II system from different sources, as exemplified for the 21 ACPs presented in the dendrogram in Fig. 7. The *eryA* ACPs are closely related to each other, except for ACP-S, which, as described above, does not function as all other known ACPs and is less related to the other SU ACPs than the ACP from 6msas. Nonetheless, the Type-I PKS ACPs appear to be clustered together indicating greater overall similarity amongst each other than with ACPs from other systems. Similarly, the Type-II PKS ACPs form their own cluster as do both the Type-I and the Type-II FAS ACPs (Fig. 7). The determination that the SU ACPs more closely resemble Type-I FAS systems than the monofunctional FAS ACP from the same host suggests that Type-I PKSs and Type-I FASs had a common ancestor. This hypothesis is corroborated by the observation of a similar pattern when the six *eryA* KSs were compared

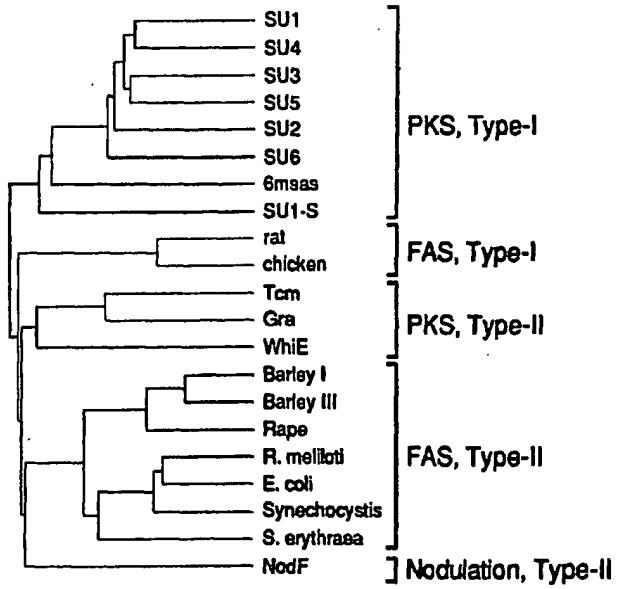


Fig. 7. Relatedness of ACPs and ACP domains. The ACP domains from multifunctional FAS and PKS systems (see Fig. 2) and the monofunctional ACPs are shown as the dendrogram obtained from PILEUP. ACP sequences: WhiE, spore-associated pigment genes from *Streptomyces coelicolor* (Davis and Chater, 1990); Barley I and III, forms I and III (Hansen, 1987); Rape, *Brassica napus* plastid seed (Safford et al., 1988); R. meliloti, *Rhizobium meliloti* constitutive ACP (Platt et al., 1990); E. coli, *E. coli* FAS ACP (Holak et al., 1988); Synechocystis, *Synechocystis* 6803 (Froehlich et al., 1990); S. erythraea, *S. erythraea* putative FAS ACP (Revill and Leadlay, 1991); NodF, *R. meliloti* nodulation-specific ACP (Debelle and Sharma, 1986).

with eight other Type-I and Type-II sequences (data not shown).

The finding of a stretch of the DH-ER interdomain region in the SU lacking these two functions is also consistent with the hypothesis that the *eryA* modules are likely to have evolved from an ancestral element (FAS- or PKS-like) which encoded the full set of activities involved in the processing of the β-carbonyl (DH, ER and KR), followed by loss of the functions not required at particular steps of 6dEB synthesis. Two modes of specialization through loss of function seem to have occurred in the *eryA* modules: selected mutations in the KR-encoding domain in module 3, and loss of the DH- and ER-encoding segments in all of the modules except module 4. Loss of function (ER) through extensive deletion may have also taken place in 6msas. It will be interesting to analyze the sequences of other PKS systems lacking KR, DH or ER domains to better understand the mode of evolution of pathways for complex polyketides.

(f) Conclusions

Our results on the extent of the various domains in the six *eryA* SU, determined solely by computer-assisted alignments, can be extended to other related systems and are substantially in agreement with those independently found by P.F. Leadlay and colleagues (personal communication) and by Witkowski et al. (1991a), who corroborated their computer analysis with limited proteolysis studies. The existence of multiple sequences with identical function in *eryA* has greatly facilitated assignments of the various domains. We have proposed a location for the DH domain and a putative active-site His for it. Type-I FAS and PKS systems also seem to share a common origin independent of their prokaryotic or eukaryotic source.

ACKNOWLEDGEMENTS

We thank Dick Hutchinson, Peter Leadlay, Stuart Smith and Rich Summers for providing us with results prior to publication.

REFERENCES

Amy, C.M., Witkowski, A., Naggert, J., Williams, B., Randhawa, Z. and Smith, S.: Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase. Proc. Natl. Acad. Sci. USA 86 (1989) 3114–3118.

Beck, J., Ripka, S., Siegner, A., Schiltz, E. and Schweizer, E.: The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. Eur. J. Biochem. 192 (1990) 487–498.

Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F. and Hutchinson, C.R.: Analysis of the nucleotide sequence of the *Streptomyces glucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. EMBO J. 8 (1989) 2727–2736.

Bloch, K.: β-Hydroxydecanoyl thioester dehydrase. In: Boyer, P.D. (Ed.), The Enzymes, 3rd ed., Vol. V. Academic Press, New York, 1971, pp. 441–464.

Chirala, S.S., Kuziora, M.A., Spector, D.M. and Wakil, S.J.: Complementation of mutations and nucleotide sequence of FAS1 gene encoding the β-subunit of yeast fatty acid synthase. J. Biol. Chem. 262 (1987) 4321–4340.

Cronan Jr., J.E., Li, W.-B., Coleman, R., Narasihman, M., de Mendoza, D. and Schwab, J.M.: Derived amino acid sequence and identification of the active site residues of *Escherichia coli* β-hydroxydecanoyl thioester dehydrase. J. Biol. Chem. 263 (1988) 4641–4646.

Cortes, J., Haydock, S.H., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F.: An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature 348 (1990) 176–178.

Davis, N.K. and Chater, K.F.: Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. Mol. Microbiol. 4 (1990) 1679–1691.

Debelle, S. and Sharma, S.B.: Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity of nodulation. Nucleic Acids Res. 14 (1986) 7453–7472.

Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 (1984) 387–395.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L.: Modular organization of genes required for complex polyketide biosynthesis. Science 252 (1991) 675–679.

Froehlich, J.E., Poorman, R., Reardon, E., Barnum, S.R. and Jaworski, J.G.: Purification and characterization of acyl carrier protein from two cyanobacteria species. Eur. J. Biochem. 193 (1990) 817–825.

Hallam, S.E., Malpartida, F. and Hopwood, D.A.: Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor*. Gene 74 (1988) 305–320.

Hansen, L.: Three cDNA clones for barley leaf acyl carrier protein I and III. Carlsberg Res. Commun. 52 (1987) 381–392.

Hess, G.P.: Chymotrypsin – chemical properties and catalysis. In: Boyer, P.D. (Ed.), The Enzymes, 3rd ed., Vol. III. Academic Press, New York, 1971, pp. 213–248.

Holak, T.A., Nilges, M., Prestegard, J.H., Gronenborg, A.M. and Clore, G.M.: Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry. Eur. J. Biochem. 175 (1988) 9–15.

Holzer, K., Liu, W. and Hammes, G.H.: Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA. Proc. Natl. Acad. Sci. USA 86 (1989) 4387–4391.

Linial, M., Miller, K. and Scheller, R.H.: VAT-1. An abundant membrane protein from torpedo colinergic synaptic vesicles. Neuron 2 (1989) 1265–1273.

McCarthy, A.D. and Hardie, D.G.: Fatty acid synthase – an example of protein evolution by gene fusion. Trends Biochem. Sci. 9 (1984) 60–63.

Piatigorsky, J. and Wistow, G.: The recruitment of crystallins: new functions precede gene duplication. Science 252 (1991) 1078–1079.

Platt, M.W., Miller, K.J., Lane, W.S. and Kennedy, E.P.: Isolation and characterization of the constitutive acyl carrier protein from *Rhizobium meliloti*. J. Bacteriol. 172 (1990) 5440–5444.

Poulose, A.J., Rogers, L., Cheesbrough, T.M. and Kolattukudy, P.E.: Cloning and sequencing of the cDNA for S-acyl fatty acid synthase thioesterase from the uropygial gland of mallard duck. J. Biol. Chem. 260 (1985) 15953–15958.

Randhawa, Z.I. and Smith, S.: Complete amino acid sequence of the medium-chain *S*-acyl fatty acid synthetase thio ester hydrolase from rat mammary gland. Biochemistry 26 (1987) 1365–1373.

Revill, W.P. and Leadlay, P.F.: Cloning, characterization and high level expression in *Escherichia coli* of the *Saccharopolyspora erythraea* gene encoding an acyl carrier protein potentially involved in fatty acid biosynthesis. J. Bacteriol. 173 (1991) 4379–4385.

Rodokanaki, A., Holmes, R.K. and Borras, T.: Zeta crystallin, a novel protein from guinea pig lens is related to alcohol dehydrogenase. Gene 78 (1989) 215–224.

Safford, R., de Silva, J., Lucas, C., Windust, J.H.C., Shedden, J., James, C.M., Sidebottom, C.M., Slabas, A.R., Tombs, M.P. and Huges, S.G.: Molecular cloning and sequence analysis of complementary DNA encoding rat mammary gland medium-chain *S*-acyl fatty acid synthetase thio ester hydrolase. Biochemistry 26 (1987) 1358–1364.

Safford, R., Windust, J.H.C., Lucas, C., de Silva, J., James, C.M., Hellyer, A., Smith, C.G., Slabas, A.R. and Huges, S.G.: Plastid-localised seed acyl-carrier protein of *Brassica napus* is encoded by a distinct, nuclear multigene family. Eur. J. Biochem. 174 (1988) 287–295.

Schweizer, M., Roberts, L.M., Holtke, H.-J., Takabayashi, K., Hollerer, E., Hoffmann, B., Muller, G., Kottig, H. and Schweizer, E.: The pentafunctional FAS1 gene of yeast: its nucleotide sequence and order of catalytic domains. Mol. Gen. Genet. 203 (1986) 479–486.

Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.H., Bibb, M.J. and Hopwood, D.A.: Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. EMBO J. 8 (1989) 2717–2725.

Tsukamoto, Y. and Wakil, S.J.: Isolation and mapping of the β-hydroxyacyl dehydratase activity of chicken liver fatty acid synthase. J. Biol. Chem. 263 (1988) 16225–16229.

Wakil, S.J.: Fatty acid synthase, a proficient multifunctional enzyme. Biochemistry 28 (1989) 4523–4530.

Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B. and McAlpine, J.B.: An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. Science 252 (1991) 114–117.

Witkowski, A., Rangan, V.S., Randhawa, Z.I., Amy, C.M. and Smith, S.: Structural organization of the multifunctional animal fatty-acid synthase. Eur. J. Biochem. 198 (1991a) 571–579.

Witkowski, A., Naggert, J., Wessa, B. and Smith, S.: A catalytic role for histidine-237 in rat mammary gland thioesterase II. J. Biol. Chem. 266 (1991b) 18514–18519.

Wong, A., Kean, L. and Maurer, R.: Sequence of the *dnaB* gene of *Salmonella typhimurium*. J. Bacteriol. 170 (1988) 2668–2675.

Yuan, Z., Liu, W. and Hammes, G.H.: Molecular cloning and sequencing of DNA complementary to chicken liver fatty acid synthase mRNA. Proc. Natl. Acad. Sci. USA 85 (1988) 6328–6331.

(a) INTRA-POLYPEPTIDE LINKER

RAL

M2ery: GGATGAEQAAPATT..APVD
M4ery: VGDAD..QAA.VRVVGAA.DES
M6ery: VGAAEAEQA.PALVREVPKDAD
M2rif: FGSA.A.NR.PAEIGTAAAE
M3rif: LG..ER.PAAPAPVTRDVSD
M5rif: GETVAGAPATPVTTVADAG
M3rap: .ELFTGENPAPVRGPVSAVGQD
M4rap: .ELFTGENPAPVRGPVSVVGQD
M7rap: .ELFTGENPAPVRGPVSA.GQD

ERL (b) N-TERMINAL INTER-POLYPETIDE LINKER

M3ery: ......VTD SE KVAEYLRR .ATLDLRAAR QRIRE..LES
M5ery: MSGDNGM.TE E.KLRRYLKR TVT.ELDSVT ARLRE..VEH RAG
M4rif: ......MSAPNE QIVDAL.R ASLKE....N VRLQQENSAL AAAAA
M7rif: ......VSASYE KVVEAL.R KSLEE....V GTLKKRNRQL ADAAG
M8rif: ......V.AD EGQLRDYLKR .AIADARDAR TRLRE..VEE QAR
M9rif: ......MATD E.KLLKYLKR .VTAELHS.. ..LRKQGARH .AD
M5rap: ......MR.. EDQLLDAL.R KSVKE....N ARLRKANTSL RAAMD
M11rap: ......M.PEQD KVVEYL.R WATAELHTTR AKL.....EA LAAANT

Figure 3

Exhibit B

Erythromycin PKS and the Rapamycin PKS:
N-Termini of PKS Proteins With N-Terminal KS Domains

|----------KS n-terminus----------

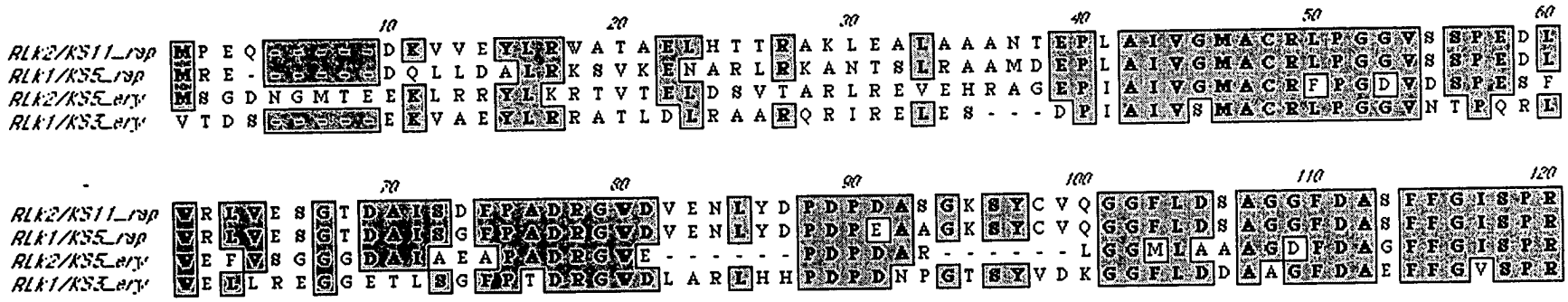


Exhibit C-1

Erythromycin PKS and the Rapamycin PKS:
C-Termini of PKS Proteins With C-Terminal ACP Domains

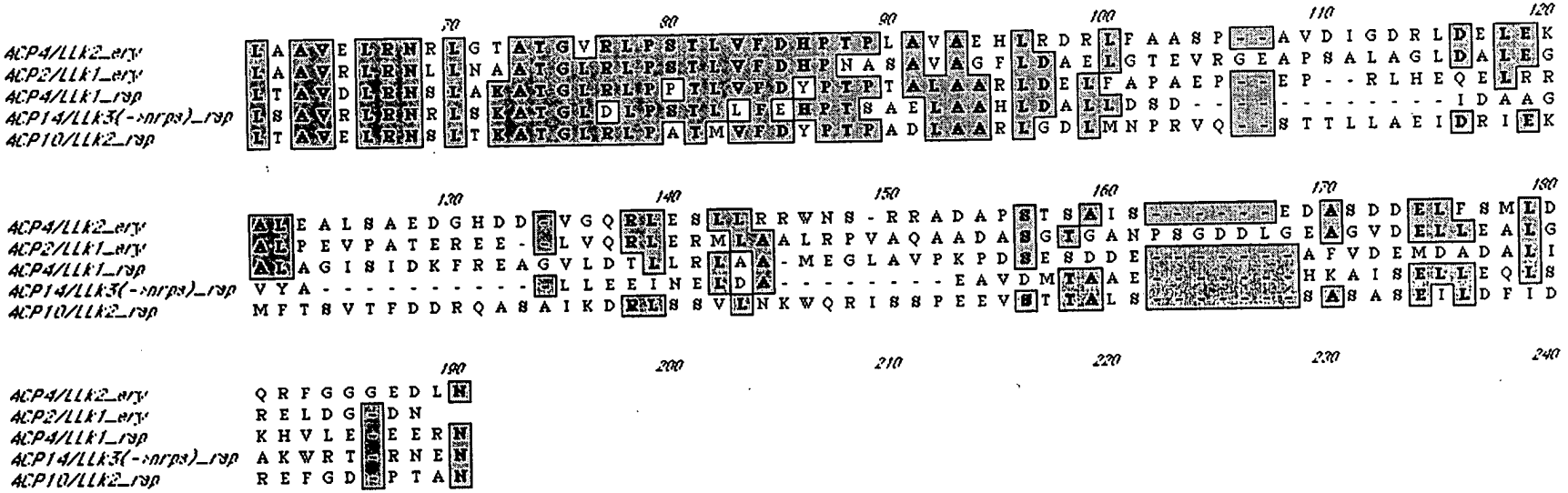


Exhibit C-2

Cter Edge of ACP (last domain of module)
& Nter edge of KS (1st domain of next module)
-all such examples from the rapamycin PKS & the erythromycin PKS

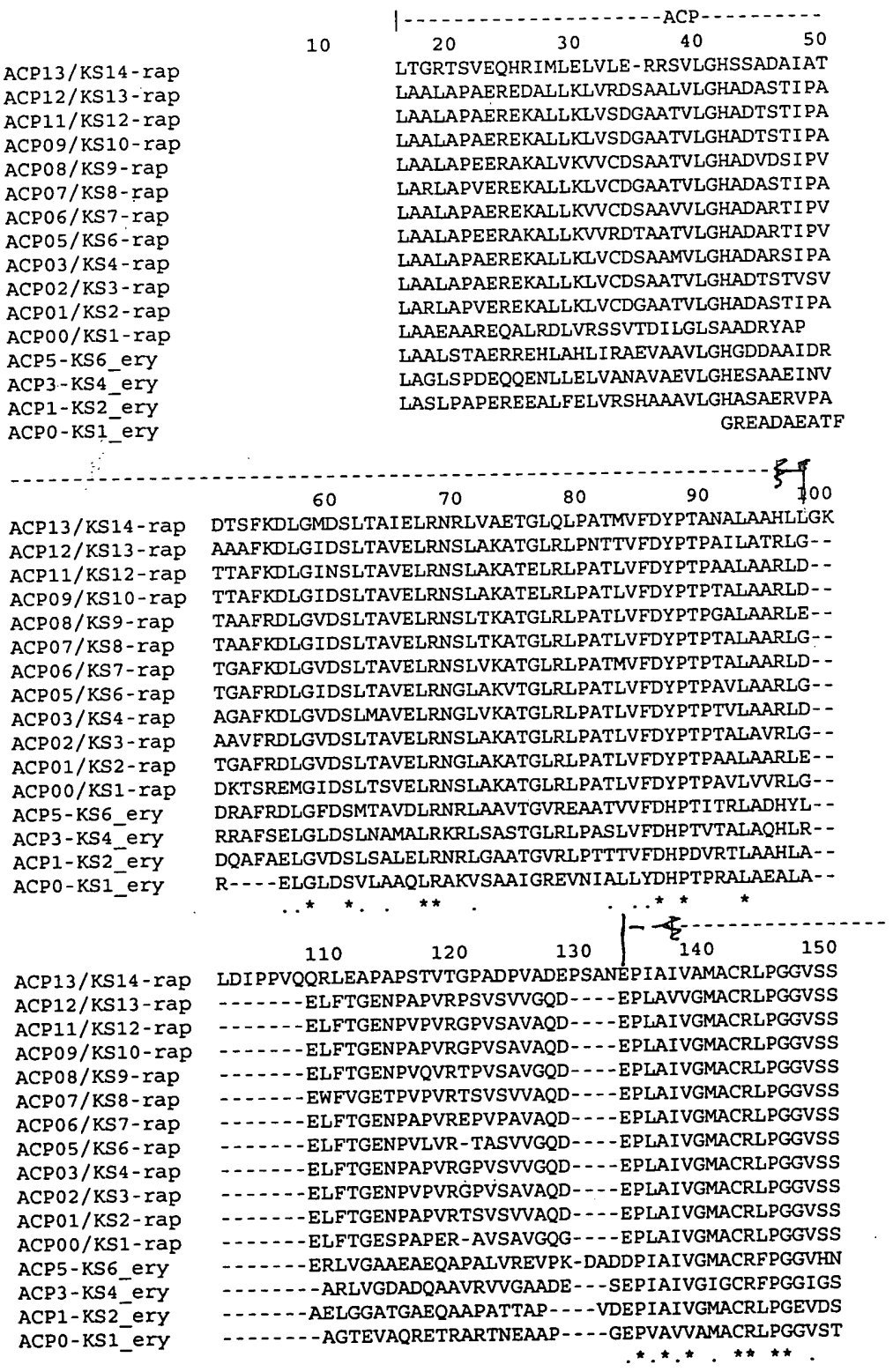


Exhibit D-1

```
-------------------------KS (N-terminus)------------------------------
                        160       170       180       190       200
ACP13/KS14-rap  PEGLWHLVESGTDAISGFPTDRGWDVEGLFDPDPDAAGKSYCVQGGFLDT
ACP12/KS13-rap  PEDLWRLVESGTDAISGFPADRGWDAESLFDPDPDASGKSYCVEGGFLDS
ACP11/KS12-rap  PEDLWRLLESGTDAVSGFPTDRGWDVENLY----DMAGKSHRAEGGFLDA
ACP09/KS10-rap  PEDLWRLVESGTDAISGFPTDRGWDVENLYDPDPDAPGKSYSVQGGFLDA
ACP08/KS9-rap   PEDLWRLVESGTDAISGFPTDRGWDVENLFDSDPDAAGKSYCVEGGFLAT
ACP07/KS8-rap   PEDLWRLLESGTDAVSGFPTDRGWDVENLFG---PAAGDSYRLQGGFLDA
ACP06/KS7-rap   PEDLWRLVESGTDAVSGFPTDRGWDVEGLFDPDPDAAGKSYRAEGGFLDT
ACP05/KS6-rap   PEDLWRLVESGTDAISGFPADRGWDAESLFDPDPDAVGKSYCVEGGFLDS
ACP03/KS4-rap   PEDLWRLVESGTDAVSGFPTDRGWDVENLYDSDPEAAGKSYCVQGGFLDT
ACP02/KS3-rap   PEDLWRLLESGTDAVSGFPTDRGWDVENLYD----MAGKSHRAEGGFLDA
ACP01/KS2-rap   PEDLWRLLESGTDAVSGFPTDRGWDVENLFG---PAVGNSYRLQGGFLDA
ACP00/KS1-rap   PEDLWRLVESGTDAISGFPTDRGWDVDGLFDPDPDASGKSYCVQGGFLDT
ACP5-KS6_ery    PGELWEFIVGRGDAVTEMPTDRGWDLDALFDPDPQRHGTSYSRHGAFLDG
ACP3-KS4_ery    PEQLWRVLAEGANLTTGFPADRGWDIGRLYHPDPDNPGTSYVDKGGFLTD
ACP1-KS2_ery    PERLWELITSGRDSAAEVPDDRGWVPDELMASD----AAGTRAHGNFMAG
ACP0-KS1_ery    PEEFWELLSEGRDAVAGLPTDRGWDLDSLFHPDPTRSGTAHQRGGGFLTE
                *   *  .      .  * ****    *                 * *.

-------------------------------------------------------------------
                        210       220       230       240       250
ACP13/KS14-rap  AADFDAPFFGISPREALGMDPQQRLLLETTWEAIERAQIDPKSLRGRDVG
ACP12/KS13-rap  AGSFDAGFFGISPREALAMDPQQRLIMEVSWEAFERAGIEPGSVRG-THR
ACP11/KS12-rap  AAGFDAGFFGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGSDTG
ACP09/KS10-rap  AAGFDASFFGISPREALAMDPQQRLMLEVSWEAFERAGIEPGSVRGSDTG
ACP08/KS9-rap   AANFDASFFGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGSDTG
ACP07/KS8-rap   AAGFDASFFGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGTDTG
ACP06/KS7-rap   AAGFDAGFFGISPREALAMDPQQRLLLEVSWEAFERAGIEPGSVRGSDTG
ACP05/KS6-rap   AASFDAGFFGISPREALAMDPQQRLIMEVSWEAFERAGIEPGSVRGSDTG
ACP03/KS4-rap   AAGFDAGFFGISPREALAMDPQQRLLLEVSWEAFERAGIEPGSVRGSDTG
ACP02/KS3-rap   AAGFDAGFFGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGSDTG
ACP01/KS2-rap   AAGFDASFFGISPREALAMDPQQRLVLEVSWEAFERAGIKPGSVRGTDTG
ACP00/KS1-rap   AAGFDASFFGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGSDTG
ACP5-KS6_ery    AADFDAAFFGISPREALAMDPQQRQVLETTWELFENAGIDPHSLRGSDTG
ACP3-KS4_ery    AADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGIDPDALRGTDTG
ACP1-KS2_ery    AGDFDAAFFGISPREALAMDPQQRQALETTWEALESAGIPPETLRGSDTG
ACP0-KS1_ery    ATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLERAGIPPTSLQASPTG
                *  **  ***..***** .*****  .* .**  * * * * ...

-------------------------------------------------------------------
                        260       270       280       290       300
ACP13/KS14-rap  VYVGGAAQGYGVGVDQQ----HDNGITGSSVSLLSGRVSYALGLEGPGVT
ACP12/KS13-rap  RLHGRVRGGYGAGADL-----GGFAATASATSVLSGRVSYFFGLEGPAIT
ACP11/KS12-rap  VFMGAYPGGYGIGADL-----GGFGATASSVSVLSGRVSYFFGLEGPAFT
ACP09/KS10-rap  VFIGAYPGGYGIGADL-----GGFGTTAGAASVLSGRVSYFFGLEGPAFT
ACP08/KS9-rap   VFMGAFPGGYGIGADL-----EGYGATA-GLNVLSGRLSYFFGLEGPAVT
ACP07/KS8-rap   VFMGAYPGGYGIGADL-----GGFGATASAVSVLSGRVSYFFGLEGPAIT
ACP06/KS7-rap   VFIGAFPVGYGAGAAR-----EGYGATA-APNVLSGRLSYFFGLEGPAIT
ACP05/KS6-rap   VFMGAYAGGYGAGADL-----GGFAATASATSVLSGRVSYFFGLEGPAIT
ACP03/KS4-rap   VFIGAFPVGYGAGFDR-----EGYGATS-GPSVLSGRVSYVFGLEGPAIT
ACP02/KS3-rap   VFMGAYPGGYGAGADL-----GGFAATASATSVLSGRVSYFFGLEGPAFT
ACP01/KS2-rap   VFMGAYPGGYGIGADL-----GGFGTTAGAVSVLSGRVSYFFGFEGPAFT
ACP00/KS1-rap   VFMGGFPGGYGAGADL-----EGFGATAGAASVLSGRVSYFFGLEGPAIT
ACP5-KS6_ery    VFLGAAYQGYGQDAVVPED-SEGYLLTGNSSAVVSGRVAYVLGLEGPAVT
ACP3-KS4_ery    VFVGMNGQSYMQLLAGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPALT
ACP1-KS2_ery    VFVGMSHQGYATGRPRPEDGVDGYLLTGNTASVASGRIAYVLGLEGPALT
ACP0-KS1_ery    VFVGLIPQEYGPRLAEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAIS
                   *     *                          . ***..*  * ***  .
```

Exhibit D-2

```
                active site Cys
                      |  310        320        330        340        350
ACP13/KS14-rap  VDTACSSSLVALHLASQALRQRECSLALVSGVSVMSSPAMFVEFSRQRGL
ACP12/KS13-rap  VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMATPQSFVEFSRQRGL
ACP11/KS12-rap  VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMATPQTFVEFSRQGGL
ACP09/KS10-rap  VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMPTPQTFVEFSRQRGL
ACP08/KS9-rap   VDTACSSSLVALHQAGYALRQGECSLALIGGVTVMATPHTFVEFSRQRGL
ACP07/KS8-rap   VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMATPQTFVEFARQGGL
ACP06/KS7-rap   MDTACSSSLVALHLAAQALRNGECSMALAGGVTVMATPEVFTEFARQRGL
ACP05/KS6-rap   VDTACSSSLVALYQAGYALRQGECSLALVGGVTVMATPQSFVEFSRKSGL
ACP03/KS4-rap   MDTACSSSLVALHLAAQALRNGECSMALAGGVTVMATPEVFTEFARQRGL
ACP02/KS3-rap   VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMATPELFTEFSRQRGL
ACP01/KS2-rap   VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMATPQTFVEFARQGGL
ACP00/KS1-rap   VDTACSSSLVALHQAGYALRQGECSLALVGGVTVMPTPQSFVEFSRQRGL
ACP5-KS6_ery    VDTACSSSLVALHSACGSLRDGDCGLAVAGGVSVMAGPEVFTEFSRQGGL
ACP3-KS4_ery    VDTACSSSLVGIHLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGL
ACP1-KS2_ery    VDTACSSSLVALHTACGSLRDGDCGLAVAGGVSVMAGPEVFTEFSRQGAL
ACP0-KS1_ery    VDTACSSSLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSL
                 .********* .  *  .**  .. .*.  **.**  *    .*.    *
```

Exhibit D-3